Detection and Analysis of Crime Patterns Using Apriori Algorithm

Arvind Venkatesh[1], Hemanth S[2], Juyin Shafaq Imtiaz Inamdar[3], Lokesh S[4]

[1,2,3]*Department of Computer Science and Engineering, The National Institute Of Engineering, Mysore*

[4]*Associate Professor, Department of Computer Science and Engineering, The National Institute Of Engineering, Mysore*

---***---

Abstract - *Data mining is a very powerful tool that is widely used in many industries. In this paper we use data mining technique to analyze and predict the crime patterns using apriori algorithm. This paper also reviews the literatures on various data mining applications, especially applications that can solve crime using data mining techniques.*

Key words: **Crime patterns detection, Data Mining, Apriori Algorithm, Association Rule Mining**

1. INTRODUCTION

Statistics on Crime Records Bureau shows that the crime rate has significantly increased in the last few years. Due the increase in the crimes it is more necessary to have a automated crime detection and prevention system. Several studies have been conducted in the past to design an efficient system that can speed up the crime investigations. However due the changing crime pattern and growing crime data it has become more important to upgrade the crime pattern detection and prevention system using latest data mining technologies. Data mining algorithms can be used to extract hidden knowledge from huge volume of data.

Even though we cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence.The predicted results cannot be assured of 100% accuracy but the results shows that our application helps in reducing crime rate to a certain extent by providing security in crime sensitive areas. So for building such a powerful crime analytics tool we have to collect crime records and evaluate it. Unfortunately due the classification of crime files by the law enforcement officials we can access the data that is publically available on blogs, social media, RSS feeds etc.
The main challenges we are facing are:

- Since the data we collect is huge it needs to be stored and analyzed.

- The data may be incomplete and inconsistent so analysis could be very difficult.

- Limitation of getting more crime data due to classification of crime files by law enforcement authorities.

The Knowledge Discovery is a very tedious process and consists of the following steps, to convert from raw data to some form on new knowledge:

1. Data cleaning: In this step we remove noise and unnecessary data from the dataset. This step is also referred to as data cleansing.
2. Data integration: In this step we combine various data sources to form one single dataset.
3. Data selection: In this step we analyze and carefully select the relevant data from the dataset.
4. Data transformation: In this step we transform the selected data into appropriate forms for the mining procedure. This step is also referred to as data consolidation
5. Data mining: It is the most important step in which we use advanced algorithms to extract patterns that are useful for our application.
6. Pattern evaluation:In this step we identify important patterns representing knowledges.
7. Knowledge representation: In this step we, visualize the knowledge data for the user.

2. RELATED WORK

There are number of papers which we have analyzed in order to determine the technology used and data mining

techniques. The review of literature will involve efficient and usable techniques such as fuzzy system and weka tool.

Some of the research paper that were reviewed during survey were: "Z-crime: A data mining tool for the detection of suspicious criminal activities based on decision tree" [1], "Association bundle- a new pattern for association analysis [2]", "Enhancing k-means algorithm for solving classification problems" [10], "Mining crime data by using new similarity measure." [4], "Crime Analysis and Prediction Using Data Mining" [6], "Improved Apriori algorithm based on Selection Criterion" [11] , "An Improvement Apriori Arithmetic based on Rough set Theory" [12].

Some of the limitations of other systems that were explored during the survey were:

1. Stores crime data and retrieves the same
2. No extraction of useful information
3. No extraction of crime patterns
4. Lack of user satisfaction
5. Less Efficient

3. PROPOSED SYSTEM

Proposed system includes modeling of crimes for finding suitable algorithms to detect the crime, precise detection, data preparation and transformation, and processing time. Proposed system identifies crime behavior, crime predicting, precise detection, and managing large volumes of data obtained from various sources. Proposed system is an automation for complaints registration, crime pattern prediction based on the previous crime details collected from various sources.

3.1 Scope and Objectives

Proposed system is a web enabled application.

1. Proposed system is an innovative crime detection system.
2. Proposed system major objective is to predict the crimes and their patterns.
3. Proposed system makes use of data mining technique called as "Association Rules" for crime pattern discovery.
4. Proposed system is an automation for early crimes pattern prediction which avoids crimes.

5. Proposed system makes use of previous crime details , date and location for the prediction of crime patterns.
6. Proposed system is an real time application where complaints are registers and crime details.

The main module is Crime Trendz Prediction Module. In this module, system predicts crime patterns based on the past crime details collected from various sources. Here system uses "Association Rule Mining" to analyze previous crime data and to extract the crime trends.

3.2 Apriori Algorithm

STEP 1: Scan the opinion data set and determine the support(s) of each item.

STEP 2: Generate L1 (Frequent one item set).

STEP 3: Use Lk-1, join Lk-1 to generate the set of candidate k - item set.

STEP 4: Scan the candidate k item set and generate the support of each candidate k – item set.

STEP 5: Add to frequent item set, until C=Null Set.

STEP 6: For each item in the frequent item set generate all non empty subsets.

STEP 7: For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence .Then add to Strong Association Rule.

For example let's consider crime types A, B, C, D and E. We represent the crime in the table 1.

Table – 1: Crime Data

ID	Crimes
1	A,C,D
2	A,C,E
3	A,B,C,E
4	B,E

The sets C1, C2, C3, L1, L2, and L3 are calculated as following using the rule Minimum confidence is 80% and Minimum Support is 50%. Finally the frequent item set (L) and Gained confidence set (G) is also obtained. Using all the

above tables the association rule is calculated.

Table -2: Set C1

Items	Support

Table -3: Set L1

Items	Support
A	75%
B	50%
C	75%
E	75%

Table -4: Set C2

Items	Support
AB	25%
AC	75%
AE	50%
BC	25%
BE	50%
CE	50%

Table -5: Set L2

Items	Support
AC	75%
AE	50%
BE	50%
CE	50%

Table -6: Set C3

Items	Support
ACE	50%
ABC	25%
ABE	25%
BCE	25%

Table -7: Set L3

Items	Support
ACE	50%

Table -8: Frequent Item Set (L)

Items	Support
A	75%
B	50%
C	75%
E	75%
AC	75%
AE	50%
BE	50%
CE	50%
ACE	50%

Table -9: Generate Confidence

RULE X	RULE Y	CONFIDENCE
{A} ⇒	{C}	100%
{C} ⇒	{A}	100%
{A} ⇒	{E}	66%
{E} ⇒	{A}	66%
{B} ⇒	{E}	100%
{E} ⇒	{B}	66%
{C} ⇒	{E}	66%
{E} ⇒	{C}	66%
{A} ⇒	{CE}	66%
{C} ⇒	{AE}	66%
{E} ⇒	{AC}	66%
{CE} ⇒	{A}	100%
{AE} ⇒	{C}	100%
{AC} ⇒	{E}	66%

STRONG ASSOCIATION RULE

{B} ⇒ {E}

$$\{CE\} \Rightarrow \{A\}$$

$$\{AE\} \Rightarrow \{C\}$$

$$\{A\} \Rightarrow \{C\}$$

$$\{C\} \Rightarrow \{A\}$$

4. IMPLEMENTATION AND RESULT

The entire project was implemented using the following configuration:

1. Frame work: DOTNET (suitable technology as it requires less time for implementation of this project)

2. IDE: Visual Studio 2010 (tool used for the development of .NET applications)

3. Front end: ASP.NET 4.0 (technology which is used for the development of internet-working applications)

4. Programming Language :
C#.NET (programming language which is used for processing purpose)

5. Back End – MS SQL Server (used to store the crime data)

Several data sources were used to gather the required data and uploaded into the database. After testing each module of the software the results were noted down. The sample data and results are shown below in the figure 1.1

Figure 1: Result of sample dataset

5. CONCLUSION

Crime are characterized which change over time and increase continuously. The changing and increasing of crime lead to the issues of understanding the crime behaviour, crime predicting, precise detection, and managing large volumes of data obtained from various sources. Research interests have tried to solve these issues. However, these researches are still gaps in the crime detection accuracy. This leads to the challenges in the field of crime detection. The challenges include modeling of crimes for finding suitable algorithms to detect the crime, precise detection, data preparation and transformation, and processing time.

6. FUTURE ENHANCEMENT

For future enhancements we can add:
- Online Complaint Registration (Public) portal to collect the data.
- Query Module (public can post queries to administrator)

7. REFERENCES

[1] Mugdha Sharma . " Z - CRIME: A data mining tool for the detection of suspicious criminal activities based on decision tree". IEEE 2014

[2] Wenxue Huang; Milorad Krneta; Limin Lin; Jianhong Wu . Association Bundle - "A New Pattern for Association Analysis". IEEE 2007

[3] Arit Thammano; Pannee Kesisung. "Enhancing K-means algorithm for solving classification problems". IEEE 2013

[4] Guangzhu Yu; Shihuang Shao; Bing Luo . "Mining Crime Data by Using New Similarity Measure". IEEE 2008

[5] Isuru Jayaweera; Chamath Sajeewa; Sampath Liyanage; Tharindu Wijewardane; Indika Perera; Adeesha Wijayasiri "Crime analytics: Analysis of crimes through newspaper articles". IEEE 2015

[6] Shiju Sathyadevan; Devan M. S.; Surya S Gangadharan. "Crime Analysis and Prediction Using Data Mining". First International Conference on Networks & Soft Computing (ICNSC), 2014

[7] Ubon Thongsatapornwatana, "A Survey Of Data Mining Techniques For Analyzing Crime Patterns", 2016 Second Asian Conference on Defence

Technology (ACDT)

[8] P.Dhakshinamoorthy, T.Kalaiselvan, "Crime Pattern Detection Using Data Mining", International Journal of Advanced Research in Computer Science and Applications Vol.1, Issue 1, July 2013

[9] Nidhi Tomar, Amit Kumar Manjhvar, "An Improved Optimized Clustering Technique For Crime Detection", 2016 Symposium on Colossal Data Analysis and Networking (CDAN)

[10] Arit Thammano; Pannee Kesisung; "Enhancing K-means algorithm for solving classification problems", Mechatronics and Automation (ICMA), 2013 IEEE International Conference

[11] V. Vaithiyanathan; Rajeswari Kannan; Rashmi Phalnikar; Swati Tonge; "Improved apriori algorithm based on selection criterion", Computational Intelligence & Computing Research (ICCIC), 2012 IEEE International Conference

[12] Chen, Chu-xiang & Shen, Jian-jing & Chen, Bing & Shang, Chang-xing & Wang, Yun-cheng. (2011). "An Improvement Apriori Arithmetic Based on Rough Set Theory". Proceedings - PACCS 2011: 2011 3rd Pacific-Asia Conference on Circuits, Communications and System. 10.1109/PACCS.2011.5990261.